Biological model representation and analysis

Cao, L.

Chapter 6

Phenotype analysis of arborized structures from 3D models

Extended from:

L. Cao, F.J. Verbeek. Nature inspired phenotype analysis with 3D model representation optimization. The 4th International Conference on Innovations in Bio-Inspired Computing and Applications,Ostrava, 2013

Abstract: In biology 3D models should reflect nature as close to true nature as possible that these models can be used for a accurate analysis. The visualization of these models helps in the further understanding and conveying of the research problem. In Chapter 5 we have introduced a solution of 3D model representation optimization. We verified our pipeline under different levels of complexity of 3D structures. In this Chapter, we further focus on the investigation of complex structures from lactiferous duct of newborn mice to gain understanding of the 3D structure and the deviation as a result of enviromental factors. We make use of Lindenmayer systems (L-systems) and the results of our analysis for gaining understanding of lactiferous duct development. The approach also produces a framework for analysis of complex 3D models derived from 3D images.
The optimized 3D models we obtain through the pipeline of Chapter 5 are used in the analysis of phenotypical differences originating from experimental conditions by extracting related shape features from the model. In order to make sure we can extract the branching structure in the right manner we make use of the centerline extraction method and further design an evaluation method to verify the method. We consider that the lactiferous duct has an innate blue-print of its arborazation and assume this blue-print is kind of rewriting rule as can be simulated with an innate L-system. We analyze the duct as it is exposed to different environmental conditions and reflect on the effect on the innate L-system.
Our method can deal with the complex 3D models, the features separate the experimental conditions. The results provide a means to reflect on the manipulation of an L-system through external factors.

6.1 Introduction

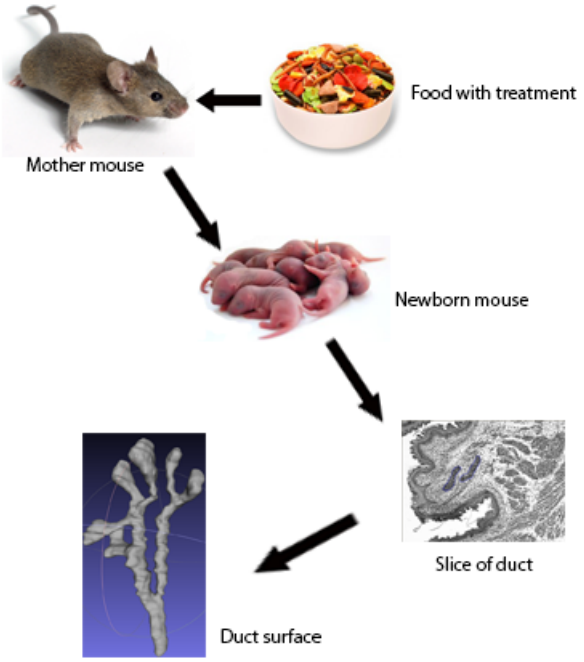

Figure 6.1. Overview of the data acquisition process.

In the previous Chapter we have illustrated the optimization of contour based reconstruction method. We make use of the method for the optimization and surface reconstruction of a complex structure: lactiferous duct of newborn mice. The study of the lactiferous duct aims to illustrate the effect of substances in the environment that mimic hormones and therefore have a potential effect on development of sexual organs or secondary sexual organs. This effect is known as endocriene disruption [Mandrup et al., 2014]. The study consists of 4 groups. In the data set presented here 4 different conditions are introduced; meaning that through the food the mother was exposed to these conditions and we would like to measure a maternal effect in the offspring. The control group: wildtype is not exposed (WT). A conditional control group is exposed to an inert component (OLIE). One group is exposed to a range of concentrations of diethylstilbestrol (DES). Another group is exposed to a cocktail of estrogen and progesterone (EP) in a range of different concentrations. An overview of the experiment is shown

in Figure 6.1. From the reconstruction it shows that the mouse lactiferous duct has distinct tree-like structure as depicted in Figure 6.2.

We want to illustrate the capability of our 3D representation method in dealing with complex structures and introduce a workflow for 3D phenotype analysis. This Chapter therefore focuses on the extraction of the complex topological features to detect the morphological changes under different treatment conditions. Graph based measurements have the potential of encoding geometrical and topological shape properties in a more faithful and intuitive manner [Akgül et al., 2009]. The skeleton is one of them defined as a thinned version of the shape of which all elements are equidistant to its boundaries. In 3D, skeletons generally contain surface patches. Centerline is a special case of the skeleton in which the skeleton is simplified to a 1D representation of the original 3D object, consisting only of curves [Dey and Sun, 2006]. In case of an arbor-like structure, the centerline is derived from the two outermost sections of an arborized structure which locally maximizes the distance from the boundary [Piccinelli et al., 2009]. The centerline describes the topology information of arborized structures in a compact and efficient way.

Numerous centerline extraction methods have been presented for medical imaging. They can be separated into three categories: topological thinning, distance-based methods and polygon-based methods.

The topological thinning deals with a volumetric dataset. Boundary voxels are iteratively peeled off according to topological rules until the whole volume has been peeled and left with the centerline [Lee et al., 1994; Xie et al., 2003]. Thinning procedures usually create centerline connected of one voxel thick. However, undesirable small branches are also created on coarse object boundaries due to the principle of end-point preservation. End-point means that the point on the one voxel graph has less than 2 neighbors. As a result, the topological thinning method needs to be coupled with an efficient pruning method [Nmeth et al., 2010].

Distance-based methods compute the boundary distance transform and find the centerline as the local maximum or ridge. However, the output centerline is usually disconnected, not guaranteed to be one voxel thick, and quite sensitive to noise on the boundary [Cardenes et al., 2010; Van Uitert and Bitter, 2007]. The reason is that the local maximum extraction might produce a surface based skele-

ton and not a centerline. There would exist a planar set of voxels in the middle of the object which would comply with the local maximum criterion [Telea and Vilanova, 2003].
The polygon-based methods directly deal with the surface and the output centerline is suitable to topological analysis of arborized structures. It is useful to obtain a centerline of a complex; e.g. arborized network. It however is a heuristic method which lacks a sound theoretical framework and guaranteed solution [Piccinelli et al., 2009]. In our case it has shown to perform well.
The centerline preserves the minimal topology of an arborized surface, therefore, for our research, the features from the centerline can be used to describe relevant parts of the arborization structure. Furthermore, several prominent features that distinguish the various treatment groups are helpful in the construction of a formalized system. To that end we investigate the connection from an arbor-like structure to an L-system [Lindenmayer, 1968; Rozenberg and Salomaa, 1980, 1986] and reason over the phenotypical effects as a result of different conditions to which the subjects have been exposed.
The L-systems have been applied successfully in arborizations such as neural tissue [Ascoli and Krichmar, 2000; Jelinek et al., 2002]. The computational construction of neurons using L-system provide an efficient and reliable method to investigate the relationships between neuromorphology and neurophysiology [Ascoli, 1999; Jelinek et al., 2004]. The mammary gland also has a noticeable branching feature; as a result, we intend to use L-system to simulate the mammary gland structure. L-systems were firstly introduced as a mathematical system to describe the developmental patterns of algae in 1968 [Lindenmayer, 1968] by a theoretical biologist: Aristide Lindenmayer. Subsequently, Lindenmayer's method was extended with a geometric interpretation and became a versatile method for plant modeling. In general, an L-system is a system of rewriting. Rewriting is a technique for defining complex objects by successively replacing parts of a simple initial object using a set of rewriting rules or productions [Prusinkiewicz and Lindenmayer, 1990]. The most known rewriting rules operate on character strings. L-systems are one of them with distinctive characteristics. Specifically, L-systems consist of two trivial elements: an axiom, and a set of productions. The axiom is the starting point of the rewriting process. The set of productions are the

derivative rules. In L-systems, the productions are applied on a parallel and simultaneous manner replacing all letters in a given string.

6.2 Methodology for arborization structure analysis

In order to understand the arborization structure we first have to extract the topological features from the optimized 3D surface model. We investigate how this can be accomplished with a minimal representation of the topology of a model i.e. the centerline. It facilitates quantitative analysis of the arborized structures. Arborization structure analysis is discussed in many areas. Researchers have recently proposed methods to trace neurons or reconstruct the neuronal structure by tracking arborized line structures in a 3D image volume [Lee et al., 2012; Zhao et al., 2011]. The quantitative analysis of the tree is substantially simplified by transforming a voxel-level tree object into a set of interconnected single-voxel centerlines representing individual tree branches [Kalman et al., 2006].
Centerline analysis is also used for the blood vessel quantification and meaningful parameter extraction from 3D vascular images [Kang et al., 2009].
All considered, we have therefore chosen the centerline extraction method. Extra and useful topology related and relevant measurements for the analysis of the lactiferous duct structures, can be extracted from the centerline; i.e. number of branches, average length of branches, number of bifurcations, and so on.

6.2.1 Centerline extraction

Polygon-based centerline extraction method can create a connected one voxel thick centerline without the undesired small branches. Futhermore, it can be implemented in fully automatic fashion which help to prevent subjective information entering into the process; i.e. the approach make the subsequent phenotype measurements more objective. The method we implement in our workflow is a mesh-based centerline extraction created by Piccinelli et al. [Piccinelli et al., 2009] for identifying vessel bifurcations and arborized network organization. First,

the algorithm identifies the bifurcations by analyzing the whole network of the arborized model with a moving sphere. In this manner the method is able to decompose the network into bifurcations and single branches. Since the complexity of the algorithm is linear with the number of surface mesh vertices, it is very efficient. There are, however, some limitations. Nested bifurcations with a very short distance between each other may be grouped into a single bifurcation if the probing sphere radius on the junction part is greater than the distance between the bifurcations [Piccinelli et al., 2009]. Centerlines can extend outside the model if the boundary has a particularly non-convex shape [Piccinelli et al., 2009]. As a result, we evaluate this centerline extraction method in section 6.3 so as to check the performance specifically, with our dataset in mind.

6.2.2 Evaluation of centerline extraction method

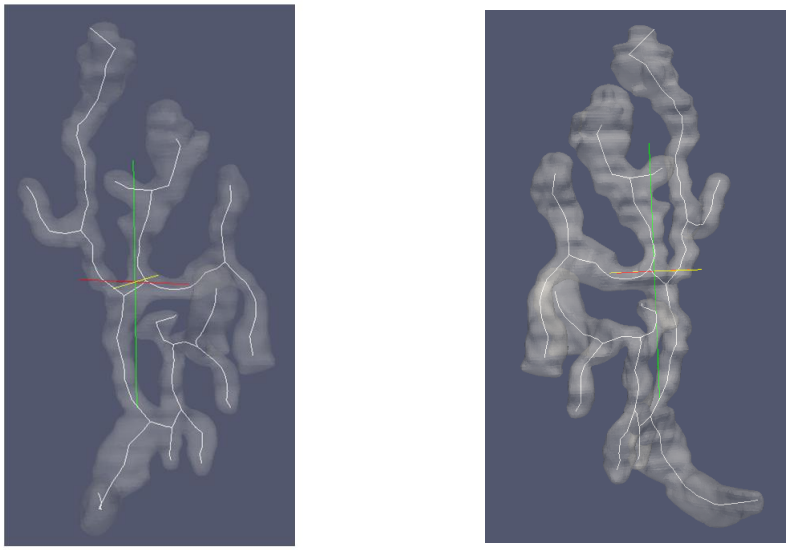

Figure 6.2. Branch structure and surface model with opacity.

Since the centerline extraction method has its limitations, it would be necessary for us to evaluate it in our own dataset. With a good evaluation, we can safely continue with the measurements, classification and modeling. The ground truth for the evaluation is the original surface model. We verify the correctness of the centerline in the ParaView software environment [Henderson, 2004] as shown in Figure 6.2. We compare each branch and bifurcation between original model and

Table 6.1. Evaluation result of centerline extraction method

Model name	#Branch	#Bifurcation	#Wrongly connected branch	#Low quality bifurcation
DES_1	6	3	0	0
DES_2	11	5	0	0
DES_3	5	2	0	0
DES_4	21	9	0	0
EP_1	11	7	1	1
EP_2	7	3	1	0
EP_3	25	12	1	0
EP_4	23	11	1	0
OLIE_1	17	8	0	1
OLIE_2	17	8	0	0
OLIE_3	5	2	0	0
OLIE_4	9	4	0	0
WT_1	7	3	0	0
WT_2	9	4	0	0
WT_3	5	2	0	0
WT_4	5	2	0	0
Total	183	85	4	2

DES: a group exposed to a range of concentration of diethylstilbestrol; EP:a group exposed to a cocktail of estrogen and progesterone; OLIE: a condition control group exposed to an inert component; WT: a control group.

cneterline so as to make sure they are extracted correctly. In [Piccinelli et al., 2009] there are two limitations for this method. One is the failed identification of two bifurcations which are too close to each other. The other is the centerlines that can extend outside the "ducts". In the mammary gland model, however, we do not find so many exceptional cases. In some cases, we find branches that are connected wrong. Some of the bifurcations locate far from the arbor center and do not nicely represent the original topology. Therefore, in an initial evaluation,

we focus on number of branches wrongly connected and the low quality bifurcations. We have randomly selected 16 out of 42 mammary gland models with automatic extraction of the centerlines and calculate the number of branches and bifurcations per model as well as the wrongly connected branches and low quality bifurcations. The results are shown in Table 6.1.
As we can compute from Table 6.1, 97.81% of the branches are detected with correct topology of the tree structure in mammary gland. Those who are connected wrong which are mostly occurring in EP condition are due to the sequence that the topology structure is relatively complicated and some small portions of the surface from two branches stick together because they are too close to each other and the interpolation method cannot separate them apart. Additionally 97.65% of the bifurcations are properly situated in the model. Although the automatic extracted centerlines do not perfectly fit to the ideal centerline, it sufficiently detects most of the branches and bifurcations correctly. Conclude that we can safely apply this automatic centerline extraction method for our model and compute features from it.

6.3 L-system construction

Within the feature set we have obtained from our experiments, we look for features that is prominent in representing phenotype changes of mammary gland under different treatment. Subsequently, we can generate a general L-system for mammary gland simulation by making use of these features.
The mammary gland that we study with is from newborn mouse whose mothers were exposed to different endocrine-like components. It is in the stage of secondary sprout: lactiferous duct or epithelial cords. In human, the amount of lactiferous ducts could reach to 10-20 for each nipple. However, in the mouse embryo, only one epithelial cord grows from the mammary bud. Thus, only one lactiferous duct exists for each nipple. This initial round of branching growth results in a primary duct with several initial branches [Cowin and Wysolmerski, 2010].
After extracting several prominent branch features from the centerline of actual 3D lactiferous duct, we would introduce these morphological features into an L-

system to model simulation of the lactiferous duct.
The platform we have chosen to construct the L-system model for lactiferous duct is L-Py: an L-system simulation framework for modeling plant architecture development based on the Python language [Boudon et al.]. We use a dynamic language to enhance the development of lactiferous duct growth models. It provides seamless control of the differential turtle geometry [Prusinkiewicz and Lindenmayer, 1990]. The turtle is a drawing tool for the geometric interpretation of a character string as a sequence of segments. During the string reading, the turtle moves on the line with varying segment lengths and angles. The turtle can move in the 3D space based on differential geometry and using quantities such as local tangent, curvature, and step size [Boudon et al.].
As for our first construction of lactiferous duct model using an L-system as described in a pseudo code of Algorithm 1, the production rules mainly focus on realizing the elongation and bifurcation of the duct. The measurements that we integrate in the system include branch length, curvature, radius and the number of branches. These features are the prominent ones based on our previous study to distinguish the morphological variation of lactiferous duct with various treatments. The way to integrate branch length, curvature and the radius of maximal inscribed sphere are relatively straightforward. The changing number of branches is realized by changing the probability of bifurcation construction. The probability density function we use is the uniform distribution.

6.4 Results

6.4.1 material and preparation

Our study is based on 3D model of the lactiferous duct of newborn mice. There were 4 different conditions introduced: control group consisting of 7 models; a condition control group as shown in Figure 6.4(c),(d) consists of 12 models; a treated group exposed to DES consists of 8 models in different increasing concentrations; another treated group exposed to EP shown in Figure 6.4(a),(b) consists of 8 models in different increasing concentrations. The distinct tree-like structure

of the wildtype mouse lactiferous duct is depicted in Figure 6.2.

Algorithm 1 Pseudo code for L-system

```
Initialize maximum number of branch order to 4
Initialize increment of radius through time to 0.0001
Initialize maximum duration of a branch relating to the branch order
Initialize branch angle with a random number from 15 to 25 weighted with the branch order
Initialize branch angle to a random number from 3 to 7
Initialize total number of iteration of the system
for time is smaller than iteration time do
    Set time to zero
    Set branch order to zero
    if time is shorter than maximum duration of a branch then
        Continue growing with a set branch angle and length
        Set time to time +1
    else
        Create bifurcation with a set roll angle around the previous branch and a set angle in between the bifurcation
        Set branch order to order +1
    end if
    Decrement radius of the branch
end for
```

The 3D models are obtained from serial sections and acquired with a dedicated bright field microscope imaging setup [Boon et al., 2000]. The images of the lactiferous duct are acquired and a stack of aligned images is used as input for the 3D model [Verbeek and Huijsmans, 1998; Verbeek et al., 1995]. For each stack the lactiferous duct structures are delineated by a specialist resulting in initial 3D models.

For all models, model optimization (cf. Chapter 5) and surface reconstruction (cf. Chapter 4) is applied. Subsequently, the centerlines of all branches in the arbor-like structure are extracted. To derive the topological structure a known centerline extraction method is used as described in [Piccinelli et al., 2009]. We use the centerlines to assess the topological shape differences in the models. These centerlines are also the basis for the L-system representation.

For each of the models from each global centerline we extract individual branches.

For each branch we calculate phenotype measurements including branch length, curvature, torsion, tortuosity, minimal radius, maximal radius, mean of the radius and median of the radius. The curvature of a branch is defined as the average curvature over all branch-points. The curvature of a point is the inverse of the radius of the local osculating circle, i.e.

$$K = \frac{1}{R} \tag{6.1}$$

where, K is the curvature of a point and R is the radius of the local osculating circle. The torsion is defined as the degree by which the osculating plane rotates along the line [Pressley, 2010]. The tortuosity is defined as the ratio between the branch length and the distance of the branch end-points. The radius is defined as the radius of maximum inscribed sphere for each point.

6.4.2 Prominent feature extraction

The main focus of this study is topological shape description for 3D geometrical models derived from microscope image. We have obtained experimental data from the development of the lactiferous duct; in early development a mouse embryo is exposed to different environmental factors. These factors a.k.a. endocrine disruptors influence development of reproductive organs in primary and secondary stages by altering the structure of the duct.

Changes in the duct configuration are considered changes in the innate layout of the duct. The innate layout can be modeled with an L-system. So, changes in the duct layout changes the innate coding of the L-system and this will result in a different structure or phenotype. We measure the phenotype in order to show the effect of the environmental factor and at the same time we can reason from the result how the L-system is affected.

Our question is whether we can accurately separate complex 3D structure using measurements. We construct a dataset by collecting all the branches with measurements from the same treatment groups. Because the measurements are not all normally distributed, we use the KS-test to check significant differences for every measurement.

The hypothesis is that DES and EP will have an effect on the development of the lactiferous duct; this effect should not be seen in the OLIE and WT group in which development should not be affected at all. We treat the OLIE and WT group as our ground truth. The DES and EP group should be very dissimilar from the OLIE and WT group. We therefore employ the KS-test to compare three groups with the null hypothesis that the compared groups are from the same continuous distribution: one between the groups DES and EP (DES-EP); one between the groups OLIE and WT (OLIE-WT); the other one between the combined groups DES with EP and OLIE with WT (DE-OW). In Table 6.1 the results are shown. The p-value refers to the significance level. If the test rejects the null hypothesis at the set significance level, the result value h is 1 or -1. Value h is 1 if the first dataset is significantly larger than the second. Value h is -1 if the first dataset is significantly smaller than the second dataset. Otherwise, value h is 0. The significance level was set to $p = 0.01$. From the results we can see the KS-test cannot reject the null hypothesis between group OLIE and WT which means these two groups are from the same continuous distribution. Group DES and group EP are almost the same except for the curvature, because the group EP has more curved branches compared to the group DES as we observed during the evaluation of centerline extraction method (cf. Chapter 5). Nevertheless, the combined group DES with EP and group OLIE with WT are significantly different in length, curvature, maximal radius, mean of the radius and median of the radius. These features are prominent ones over torsion, tortuosity and minimal radius. The results support the hypothesis and also confirm the feasibility of our 3D analysis for such complex system.

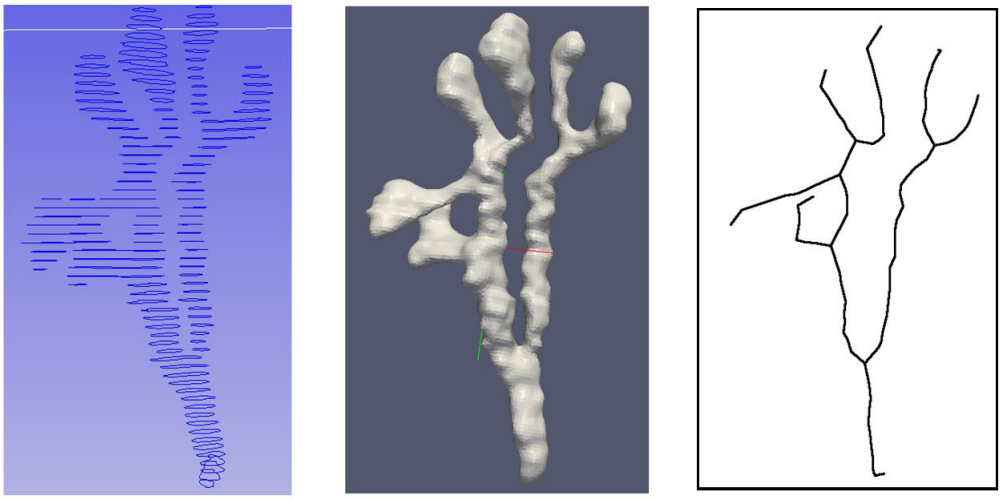

Figure 6.3. (a) original contour sections; (b) surface model; (c) centerline.

Table 6.2. Results of the different treatment groups

		length	curvature	torsion	tortuosity	min-radius	max-radius	mean-radius	median-radius
DES-EP	h	0	-1	0	0	0	0	0	0
	p	0.2027	0.0098	0.5588	0.8170	0.7951	0.6364	0.8275	0.9412
OLIE-WT	h	0	0	0	0	0	0	0	0
	p	0.2564	0.8537	0.7724	0.9246	0.0416	0.2951	0.1581	0.3943
DE-OW	h	-1	1	0	0	0	-1	-1	-1
	p	0.0043	0.0031	0.2714	0.5649	0.1007	4.99E-5	0.0026	0.0040

DES: a group exposed to a range of concentration of diethylstilbestrol; EP:a group exposed to a cocktail of estrogen and progesterone; OLIE: a condition control group exposed to an inert component; WT: a control group; DE: combined group of DES and EP; OW: combined group of OLIE and WT; h: the result of KS-test; p: the significance level of KS-test.

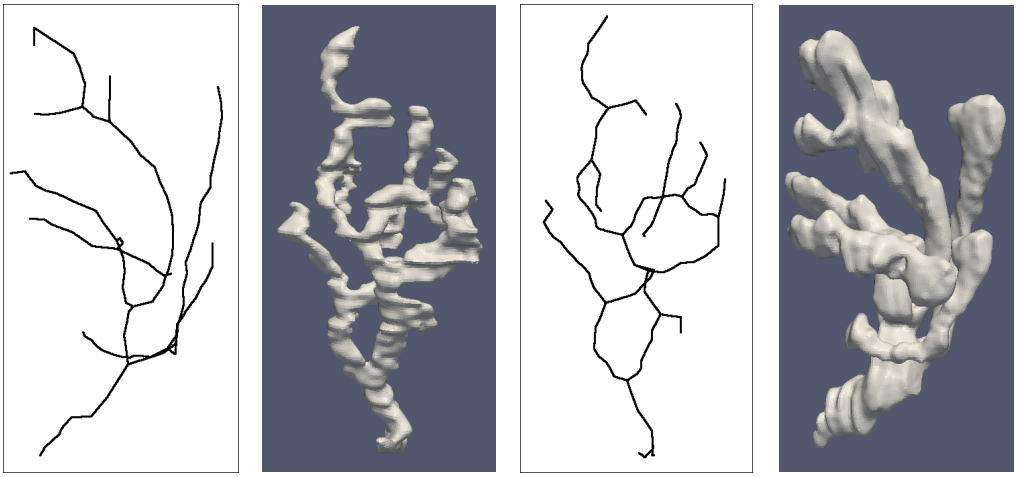

Figure 6.4. Different branch structures of lactiferous duct (a) surface model from group EP; (b) centerline of the model from group EP; (c) surface model from group OLIE; (d) centerline of the model from group OLIE.

6.4.3 L-system modeling

In this section, we intend to use the measurements, write production rules and built an L-system. This L-system model mainly focuses on the simulation of elongation and bifurcation of the centerline of the lactiferous duct. We show the simulation of our L-system model in two different kinds of lactiferous ducts: i.e. the wild type and the DES exposed type. The parameters are derived from measurement on 3D models as shown in Table 6.3.

Table 6.3. Input parameters of lactiferous duct

Type	MBL	σ_m	MBC	σ_m	MBR	σ_m	MBN	σ_m
WT	303.1	34.19	0.0096	0.0004	28.45	1.37	6.17	0.75
DES	193.1	12.85	0.0101	0.0003	27.90	0.74	14.625	3.19

WT: a control group which is not exposed; DES: a group exposed to a range of concentration of diethylstilbestrol; MBL: Mean of branch length in micrometer; MBC: Mean of branch curvature; MBR: Mean of branch radius in micrometer; MBN: Mean of branches number; σ_m: standard deviation of mean.

According to our findings, compared to wild type, the lactiferous duct with DES treatment would have significant difference such as more branches with shorter branch length. The result of our production rules and L-system visualization in

a virtual model of the lactiferous duct are shown in Figure 6.5.

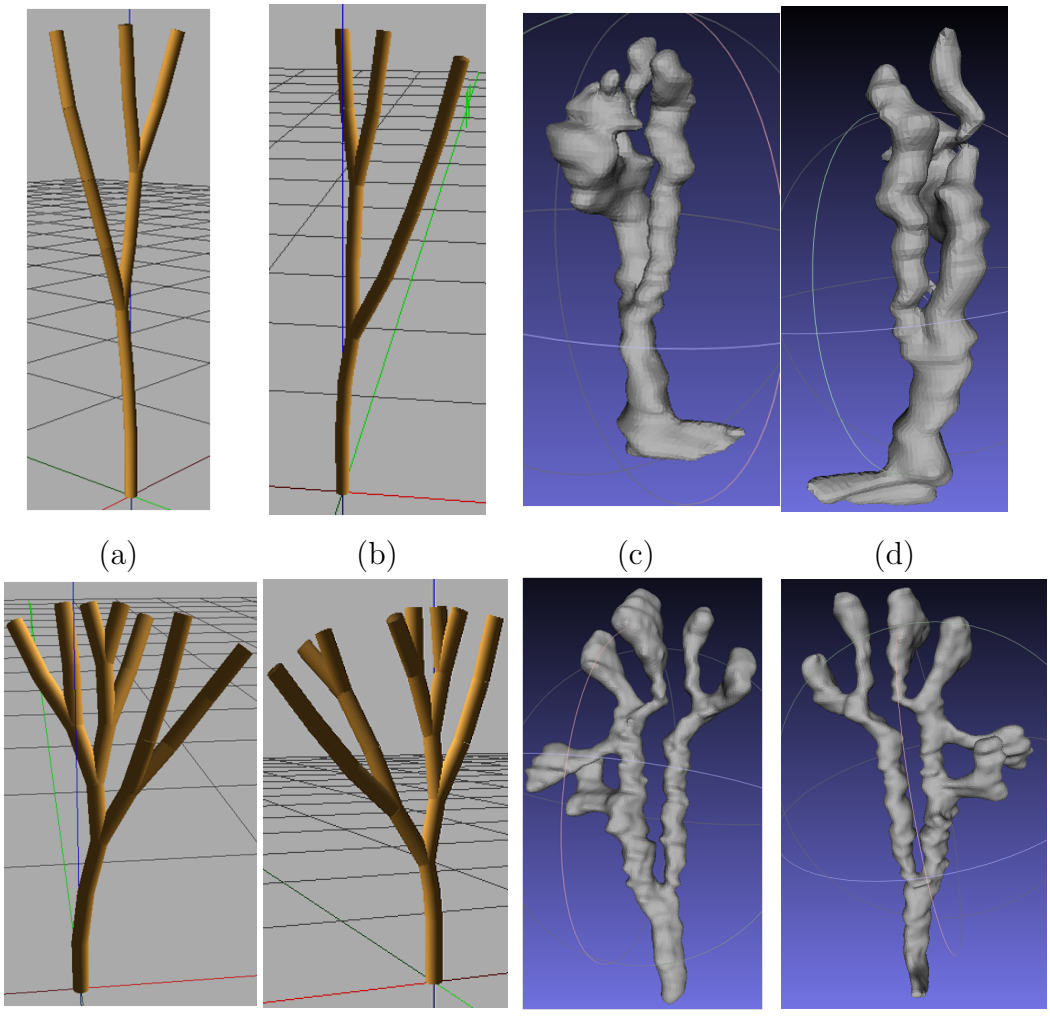

Figure 6.5. Resemblance between real surface models and L-system centerline simulation. WT centerline simulation (a) example 1 and (b) example 2; WT surface model (c) view 1 and (d) view 2; DES centerline simulation (e) example 1 and (f) example 2; DES surface model (g) view 1 and (h) view 2.

In addition, we reflect on the feasibility of ground truth of the L-system and introduce some constraints. We attempt to introduce a space constrain for the mammary gland development. Therefore, we introduce an ellipsoid as a fat pad and iterate our L-system model further to simulate the stage of puberty of the

mammary gland as shown in Figure 6.5. This is still a primitive model simulating the development of the mammary gland up to puberty stage. For a better simulation, we need extra quantified phenotype measurements to tune the model.

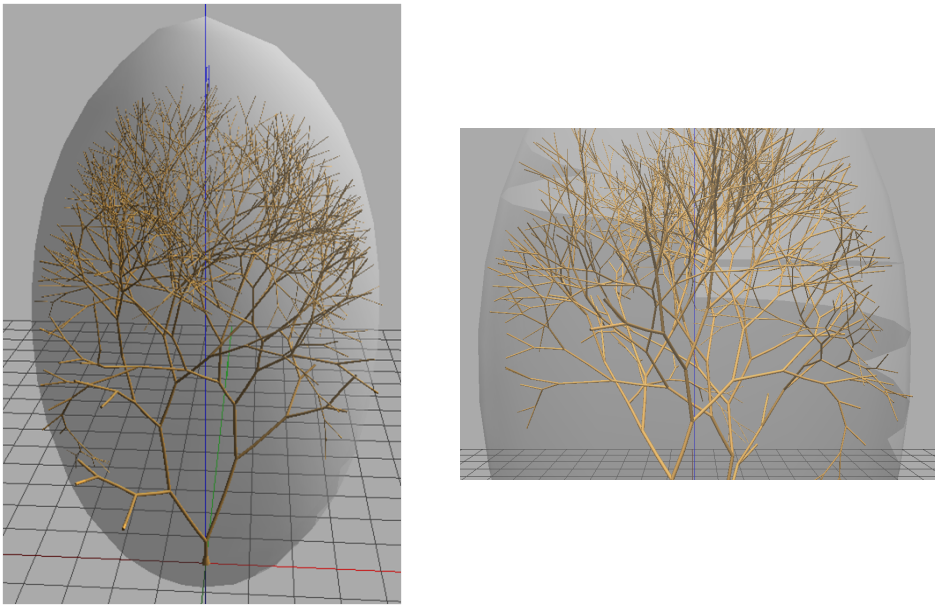

Figure 6.6. L-system simulation for the stage of puberty.

6.5 Discussion & conclusion

In this chapter, we focus on the analysis of complex 3D model with arborized structures. We study and analyze the rodent lactiferous duct under different treatment conditions. The aim is to use feature description to distinguish phenotypical changes with different treatments. We start with an optimized 3D surface model of the lactiferous duct using the method introduced in Chapter 5. We extract the centerline structure from the surface model so as to minimize the topological description. Phenotype measurements are calculated and prominent features are slected from centerline structure so as to analyse 3D models from mammary glands developed under exposure of different "potential" endocrine disruptors. We presented that we are able to differentiate the different conditions and characterize the effect of the exposures. In order to accommodate the branching structure that we analyze in a formal structure we invoke the L-system. This nature inspired formal system can help in the understanding of the branching; in

fact, we consider an L-system to represent the blue-print of the normal development of the branching duct. Our experimental results illustrate the change of the L-system under environmental conditions and the features we have derived represent the change in phenotype. At the same time the features from the various conditions represent the mutations in the L-system. Thus, this analysis of induced phenotypes helps us in bringing new inspiration to the manipulation of the L-system in which the features provide a quantifiable twist to this system. Our primary L-system model reports the potential for the simulation of mammary gland development and maturation. The L-system can provide a ground truth model from real data. Our research will be directed into this delicate interplay of nature inspired systems and nature driven models.

6.6 Acknowledgement

We wish to express our gratitude to Drs J. Lemmen and P.T. van der Saag who were involved producing the dataset for the rodent mamary gland in a range of different conditions. We thank J. Korving for taking care of the histological preparations. The imaging and reconstructions were completed using the setup of the Imaging & BioInformatics group.